

09-01-00

A

**UTILITY PATENT APPLICATION TRANSMITTAL**  
**(Large Entity)***(Only for new nonprovisional applications under 37 CFR 1.53(b))*Docket No.  
13725Total Pages in this Submission  
3**TO THE ASSISTANT COMMISSIONER FOR PATENTS**Box Patent Application  
Washington, D.C. 20231

Transmitted herewith for filing under 35 U.S.C. 111(a) and 37 C.F.R. 1.53(b) is a new utility patent application for invention entitled:

**A METHOD FOR RECOVERING 3D SCENE STRUCTURE AND CAMERA MOTION DIRECTLY FROM IMAGE INTENSITIES**

and invented by:

**John Oliensis**If a **CONTINUATION APPLICATION**, check appropriate box and supply the requisite information:☐ Continuation ☐ Divisional ☐ Continuation-in-part (CIP) of prior application No.: \_\_\_\_\_

Which is a:

☐ Continuation ☐ Divisional ☐ Continuation-in-part (CIP) of prior application No.: \_\_\_\_\_

Which is a:

☐ Continuation ☐ Divisional ☐ Continuation-in-part (CIP) of prior application No.: \_\_\_\_\_

Enclosed are:

**Application Elements**

1. ☒ Filing fee as calculated and transmitted as described below
2. ☒ Specification having 24 pages and including the following:
  - a. ☒ Descriptive Title of the Invention
  - b. ☐ Cross References to Related Applications *(if applicable)*
  - c. ☐ Statement Regarding Federally-sponsored Research/Development *(if applicable)*
  - d. ☐ Reference to Microfiche Appendix *(if applicable)*
  - e. ☒ Background of the Invention
  - f. ☒ Brief Summary of the Invention
  - g. ☒ Brief Description of the Drawings *(if drawings filed)*
  - h. ☒ Detailed Description
  - i. ☒ Claim(s) as Classified Below
  - j. ☒ Abstract of the Disclosure

**UTILITY PATENT APPLICATION TRANSMITTAL**  
**(Large Entity)**

*(Only for new nonprovisional applications under 37 CFR 1.53(b))*

Docket No.

13725

Total Pages in this Submission

3

**Application Elements (Continued)**

3. ☒ Drawing(s) *(when necessary as prescribed by 35 USC 113)*
- a. ☒ Formal                      Number of Sheets 2
- b. ☐ Informal                      Number of Sheets \_\_\_\_\_
4. ☒ Oath or Declaration
- a. ☒ Newly executed *(original or copy)*                      ☐ Unexecuted
- b. ☐ Copy from a prior application (37 CFR 1.63(d)) *(for continuation/divisional application only)*
- c. ☒ With Power of Attorney                      ☐ Without Power of Attorney
- d. ☐ DELETION OF INVENTOR(S)  
Signed statement attached deleting inventor(s) named in the prior application,  
see 37 C.F.R. 1.63(d)(2) and 1.33(b).
5. ☐ Incorporation By Reference *(usable if Box 4b is checked)*  
The entire disclosure of the prior application, from which a copy of the oath or declaration is supplied  
under Box 4b, is considered as being part of the disclosure of the accompanying application and is hereby  
incorporated by reference therein.
6. ☐ Computer Program in Microfiche *(Appendix)*
7. ☐ Nucleotide and/or Amino Acid Sequence Submission *(if applicable, all must be included)*
- a. ☐ Paper Copy
- b. ☐ Computer Readable Copy *(identical to computer copy)*
- c. ☐ Statement Verifying Identical Paper and Computer Readable Copy

**Accompanying Application Parts**

8. ☒ Assignment Papers *(cover sheet & document(s))*
9. ☐ 37 CFR 3.73(B) Statement *(when there is an assignee)*
10. ☐ English Translation Document *(if applicable)*
11. ☐ Information Disclosure Statement/PTO-1449                      ☐ Copies of IDS Citations
12. ☐ Preliminary Amendment
13. ☒ Acknowledgment postcard
14. ☒ Certificate of Mailing

☐ First Class    ☒ Express Mail *(Specify Label No.):* EL-168-497-285-US

**UTILITY PATENT APPLICATION TRANSMITTAL**  
**(Large Entity)**

(Only for new nonprovisional applications under 37 CFR 1.53(b))

Docket No.  
13725

Total Pages in this Submission  
3

**Accompanying Application Parts (Continued)**

15. ☐ Certified Copy of Priority Document(s) (if foreign priority is claimed)

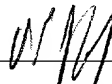
16. ☐ Additional Enclosures (please identify below):

**Fee Calculation and Transmittal**

**CLAIMS AS FILED**

For	#Filed	#Allowed	#Extra	Rate	Fee
Total Claims	13	- 20 =	0	x \$18.00	\$0.00
Indep. Claims	1	- 3 =	0	x \$78.00	\$0.00
Multiple Dependent Claims (check if applicable) <input type="checkbox"/>					\$0.00
BASIC FEE					\$690.00
OTHER FEE (specify purpose)					\$0.00
TOTAL FILING FEE					\$690.00

- ☒ A check in the amount of \$690.00 to cover the filing fee is enclosed.
- ☒ The Commissioner is hereby authorized to charge and credit Deposit Account No. 19-1013 SSMP as described below. A duplicate copy of this sheet is enclosed.
- ☐ Charge the amount of as filing fee.
- ☒ Credit any overpayment.
- ☒ Charge any additional filing fees required under 37 C.F.R. 1.16 and 1.17.
- ☐ Charge the issue fee set in 37 C.F.R. 1.18 at the mailing of the Notice of Allowance, pursuant to 37 C.F.R. 1.311(b).



Signature

Paul J. Esatto, Jr.  
Registration No. 30,749  
Scully, Scott, Murphy & Presser  
400 Garden City Plaza  
Garden City, New York 11530  
(516) 742-4343

Dated: August 31, 2000

CC:

**CERTIFICATE OF MAILING BY "EXPRESS MAIL" (37 CFR 1.10)**Applicant(s): **John Oliensis**

Docket No.

**13725**Serial No.  
unassignedFiling Date  
herewithExaminer  
unassignedGroup Art Unit  
unassigned

Invention:

**A METHOD FOR RECOVERING 3D SCENE STRUCTURE AND CAMERA MOTION DIRECTLY FROM IMAGE INTENSITIES**1c812 U.S. PTO  
09/652820

08/31/00

I hereby certify that the following correspondence:

**New Application***(Identify type of correspondence)*

is being deposited with the United States Postal Service "Express Mail Post Office to Addressee" service under 37 CFR 1.10 in an envelope addressed to: The Assistant Commissioner for Patents, Washington, D.C. 20231

**August 31, 2000***(Date)***Anneliese Eberle***(Typed or Printed Name of Person Mailing Correspondence)**(Signature of Person Mailing Correspondence)***EL - 168-497-285-US***("Express Mail" Mailing Label Number)***Note: Each paper must have its own certificate of mailing.**

SCANNED

**A METHOD FOR RECOVERING 3D SCENE STRUCTURE AND CAMERA  
MOTION DIRECTLY FROM IMAGE INTENSITIES**

**RELATED APPLICATION**

5           The present application is related to U.S. application Serial No.           , titled A  
Method for Recovering 3D Structure and Camera Motion from Points, Lines and/or  
Directly from the Image Intensities, filed on           by the same inventor as the  
present application, which related application is incorporated herein by reference.

**BACKGROUND OF THE INVENTION**

10           **1. Field of the Invention**

The present invention relates generally to a method for recovering the camera  
motion and 3D scene structure and, more particularly, to a linear algorithm for recovering  
the structure and motion directly from the image intensities where the camera moves  
along a line.

15           **2. Prior Art**

The science of rendering a 3D model from information derived from a 2D image  
predates computer graphics, having its roots in the fields of photogrammetry and  
computer vision.

Photogrammetry is based on the basic idea that when a picture is taken, the 3D  
20   world is projected in perspective onto a flat 2D image plane. As a result, a feature in the  
2D image seen at a particular point actually lies along a particular ray beginning at the  
camera and extending out to infinity. By viewing the same feature in two different  
photographs the actual location can be resolved by constraining the feature to lie on the  
intersection of two rays. This process is known as triangulation. Using this process, any

point seen in at least two images can be located in 3D. It is also possible to solve for unknown camera positions as well with a sufficient number of points. The techniques of photogrammetry and triangulation were used in such applications as creating topographic maps from aerial images. However the photogrammetry process is time intensive and inefficient.

Computer vision techniques include recovering 3D scene structure from stereo images, where correspondence between the two images is established automatically from two images via an iterative algorithm, which searches for matches between points in order to reconstruct a 3D scene. It is also possible to solve for the camera position and motion using 3D scene structure from stereo images.

Current computer techniques are focused on motion-based reconstruction and are a natural application of computer technology to the problem of inferring 3D structure (geometry) from 2D images. This is known as Structure-from-Motion. Structure from motion (SFM), the problem of reconstructing an unknown 3D scene from multiple 2D images of it, is one of the most studied problems in computer vision.

SFM algorithms are currently known that reconstruct the scene from previously computed feature correspondences, usually tracked points. Other algorithms are direct methods that reconstruct from the images' intensities without a separate stage of correspondence computation. The method of the present invention presents a direct method that is non-iterative, linear, and capable of reconstructing from arbitrarily many images. Previous direct methods were limited to a small number of images, required strong assumptions about the scene, usually planarity or employed iterative optimization and required a starting estimate.

Most SFM algorithms that are currently known reconstruct the scene from previously computed feature correspondences, usually tracked points. Other algorithms are direct methods that reconstruct from the images intensities without a separate stage of correspondence computation. Previous direct methods were limited to a small number of images, required strong assumptions about the scene, usually planarity or employed iterative optimization and required a starting estimate.

These approaches have complementary advantages and disadvantages. Usually some fraction of the image data is of such low quality that it cannot be used to determine correspondence. Feature-based methods address this problem by pre-selecting a few distinctive point or line features that are relatively easy to track, while direct methods attempt to compensate for the low quality of some of the data by exploiting the redundancy of the total data. Feature-based methods have the advantage that their input data is relatively reliable, but they neglect most of the available image information and only give sparse reconstructions of the 3D scene. Direct methods have the potential to give dense and accurate 3D reconstructions, due to their input data's redundancy, but they can be unduly affected by large errors in a fraction of the data.

A method based on tracked lines is described in "A Linear Algorithm for Point and Line Based Structure from Motion", M. Spetsakis, CVGIP 56:2 230-241, 1992, where the original linear algorithm for 13 lines in 3 images was presented. An optimization approach is disclosed in C.J. Taylor, D. Kriegmann, "Structure and Motion from Line Segments in Multiple Images, " PAMI 17:11 1021-1032, 1995. Additionally, in "A unified factorization algorithm for points, line segments and planes with uncertainty models" K. Morris and I. Kanade, ICCV 696-702, 1998, describes work on

lines in an affine framework. A projective method for lines and points is described in "Factorization methods for projective structure and motion", B. Triggs, CVPR 845-851, 1996, which involves computing the projective depths from a small number of frames. "In Defense of the Eight-Point Algorithm: PAMI 19, 580-593, 1995, Hartley presented a  
 5 full perspective approach that reconstructs from points and lines tracked over three images.

The approach described in M. Irani, "Multi-Frame Optical Flow Estimation using Subspace Constraints," ICCV 626-633, 1999 reconstructs directly from the image intensities. The essential step of Irani for recovering correspondence is a multi-frame  
 10 generalization of the optical-flow approach described in B. Lucas and T. Kanade, "An Iterative Image Registration Technique with an Application to Stereo Vision", IJCAI 674-679, 1981, which relies on a smoothness constraint and not on the rigidity constraint. Irani uses the factorization of  $D$  simply to fill out the entries of  $D$  that could not be computed initially. Irani writes the brightness constancy equation (7) in matrix form as  $\Delta$   
 15  $= -DI$ , where  $D$  tabulates the shifts  $d^i$  and  $I$  contains the intensity gradients  $\nabla I(p_n)$ . Irani notes that  $D$  has rank 6 (for a camera with known calibration), which implies that  $\Delta$  must have rank 6. To reduce the effects of noise, Irani projects the observed  $\Delta$  onto one of rank 6. Irani then applies a multi-image form of the Lucas-Kanade approach to recovering optical flow which yields a matrix equation  $DI_2 = -\Delta_2$ , where the entries of  $I_2$   
 20 are squared intensity gradients  $I_a I_b$  summed over the "smoothing" windows, and the entries of  $\Delta_2$  have the form  $I_a \Delta I$ . Due to the added Lucas-Kanade smoothing constraint, the shifts  $D$  or  $d_n^i$  can be computed as  $D = -\Delta_2 [I_2]^+$  denotes the pseudo-inverse, except in smoothing windows where the image intensity is constant in at least one direction. Using



the rank constraint on D, Irani determines additional entries of D for the windows where the intensity is constant in one direction.

Any algorithm for small, linear motion confronts the aperture problem: the fact that the data within small image windows do not suffice to determine the correspondence unless one makes prior assumptions about the scene or motion. The aperture problem makes correspondence recovery a difficult and sometimes impossible global task. To avoid this, researchers typically impose a smoothness assumption. Lucas-Kanade uses a smoothing technique to address the aperture problem.

10 **SUMMARY OF THE INVENTION**

The present invention is directed to a method for recovering 3D scene structure and camera motion from image data obtained from a multi-image sequence, wherein a reference image of the sequence is taken by a camera at a reference perspective and one or more successive images of the sequence are taken at one or more successive different perspectives by translating and/or rotating the camera, the method comprising the steps of:

- (a) determining image data shifts for each successive image with respect to the reference image; the shifts being derived from the camera translation and/or rotation from the reference perspective to the successive different perspectives;
- (b) constructing a shift data matrix that incorporates the image data shifts for each image;

(c) calculating a rank-1 factorizations from the shift data matrix using SVD, with one of the rank-1 factors being a vector corresponding to the 3D structure and the other rank-1 factor being a vector corresponding to the size of the camera motions;

(d) dividing the successive images into smoothing windows;

5 (e) recovering the direction of camera motion from the first vector corresponding to the 3D structure by solving a linear equation; and

(f) recovering the 3D structure by solving a linear equation using the recovered camera motion.

10 In accordance with the method of the present invention, step (e) includes step (e) includes:

computing a first projection matrix;

15 recovering camera rotation vectors from the shift data matrix, and the first projection matrix;

computing a second projection matrix; and

recovering the direction of camera translation using the shift data matrix, the reference image, the second projection matrix and the recovered camera rotation vectors.

20 In addition, step (f) includes recovering the 3D structure from the shift data matrix, the reference image, the recovered camera rotation vectors and the recovered direction of translation vectors.

25

The method of the present invention further includes preliminary steps of recovering the rotations of the camera between each successive image; and warping all images in the sequence toward the reference image, while neglecting the translations.

The present invention provides an algorithm for linear camera motion, where the camera moves roughly along a line, possibly with varying velocity and arbitrary rotations. The approach of the present invention applies for calibrated or uncalibrated cameras (the projective case). For specificity, we focus on the calibrated case, assuming (wlog) that the focal length is 1. The method is based on the brightness constancy equation (BCE) and thus requires the motion and image displacements to be small enough so that the intensity changes between images can be modeled by derivatives at some resolution scale.

### **BRIEF DESCRIPTION OF THE DRAWINGS**

These and other features, aspects, and advantages of the methods of the present invention will become better understood with regard to the following description, appended claims, and accompanying drawings where:

FIG. 1 schematically illustrates a hardware implementation of the present invention.

FIG. 2 is a block diagram that illustrates the method of the present invention.

### **DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENT**

#### **Definitions**

The method of the present invention assumes that the 3D structure is to be recovered from an image sequence consists of  $N_I$  images of fixed size, each with  $N_p$  pixels. Let  $p_n \equiv (x_n, y_n)^T$  give the image coordinates of the  $n$ -th pixel position. Let  $I^i$  denote the  $i$ -th image, with  $i=0, 1, \dots, N_I - 1$ , and let  $I_n^i = I^i(p_n)$  denote the image intensity at the  $n$ -th pixel position in  $I^i$ . We take  $I^0$  as the reference image. Let  $P_n$  denote the 3D point imaged at  $p_n$  in the reference image, with  $P_n \equiv (X_n, Y_n, Z_n)^T$  in the coordinate system of  $I^0$ . Let  $d_n^i$  denote the shift in image position from  $I^0$  to  $I^i$  of the 3D feature point  $P_n$ . The motion of the camera is described as its translation and rotation. Let  $T^i \equiv (T_x^i, T_y^i, T_z^i)^T$  represent the camera translation between the reference image and image  $i$ , and let  $R^i$  denote the camera rotation. In accordance with the method of the present invention we parameterize a small rotation by the rotational velocity  $\omega^i \equiv (\omega_x^i, \omega_y^i, \omega_z^i)^T$ . Let a 3D point  $P$  transform as  $P' = R(P - T)$ . Let  $p_n^i \equiv (x_n^i, y_n^i)^T \equiv p_n + d_n^i$  be the shifted position in  $I^i$  of  $p_n \in I^0$  resulting from the motion  $T^i, R_i$ .

Given a vector  $V$ , define  $[V]_2$  as the length-2 vector consisting of the first two components of  $V$ . Let  $\underline{V}$  denote the 2D image point corresponding to the 3D point  $V$ :  $\underline{V} \equiv [V]_2 / V_z$ . For a 2D vector  $v$ , define the corresponding 3D point  $\bar{v} \equiv [v^T \ 1]^T$ .  $R * v$  denotes the image point obtained from  $v$  after a rotation:  $R * v \equiv (R\bar{v})$ . Let  $\hat{v} \equiv v/|v|$ .

The three rotational flows of the camera are defined as

$$r^{(1)}(x, y), r^{(2)}(x, y), r^{(3)}(x, y) \text{ by } [r^{(1)}, r^{(2)}, r^{(3)}] \equiv \left[ \begin{pmatrix} -xy \\ -(1+y^2) \end{pmatrix}, \begin{pmatrix} 1+x^2 \\ xy \end{pmatrix}, \begin{pmatrix} -y \\ x \end{pmatrix} \right].$$

Let  $\nabla I_n = \nabla I(p_n)$  represent the (smoothed gradient of the image intensities

$I^0(p_n)$  and define  $(I_{xn}, I_{yn})^T \equiv \nabla I_n$ . Similarly, let  $\Delta I_n^i$  be the change in (smoothed)

5 intensity with respect to the reference image. With no smoothing  $\Delta I_n^i = I_n^i - I_n^0$ . Let  $\Delta$

be a  $(N_I - 1) \times N_p$  matrix with entries  $\Delta I_n^i$ .

Suppose  $V^a$  is a set of quantities indexed by the integer  $a$ . The notation  $\{V\}$  is used to

denote the vector with elements given by  $V^a$ . Let the  $(N_I - 1) \times 3$  matrices

$\bar{T} \equiv [\{T_x\} \{T_y\} \{T_z\}]$  and  $W \equiv [\{\omega_x\} \{\omega_y\} \{\omega_z\}]$  encode all translations and rotational

10 velocities for a sequence. We use the notation  $\{V\}$  to denote the vector with elements

given by the  $V^a$ .

Define the  $(N_I - 1) \times (N_I - 1)$  matrix  $C^{ii'} \equiv \delta^{ii'} + 1$ , and use

$$\left[ C^{-1/2} \right]^{ii'} = \delta^{ii'} - \left( 1 + N_I^{-1/2} \right) / (N_I - 1).$$

15

### Preliminary Analysis

Before describing the method of the present invention, we shall describe the preliminary analysis used to derive the translational and rotational flow vectors to be applied in the algorithm. For small rotations and translations, the matrix of feature-shifts

20  $d_n^i$  is approximately bilinear in the unknown  $T^i, R^i, Z_n$ . (We do not assume that the

rotations are small initially, but we can take them as small following their initial recovery and compensation.) By contracting this matrix with the  $\nabla I_n$ , we get via the brightness constancy equation (described later herein) a bilinear relation between the intensity changes  $\Delta I_n^i$  and unknowns.

5

### Derivation

The derivation of the flow vectors is described as follows. Up to noise, the

feature-shift  $d_n^i$  can be written as  $d_n^i = d_{Tn}^i + d_{Rn}^i$ , (1)

$$d_{Tn}^i = \frac{Z_n^{-1}(T_z^i p_n - [T^i]_2)}{1 - Z_n^{-1} T_z^i}, \quad d_{Rn}^i \equiv p_n^i - R^{-1} * p_n^i. \quad \text{Where } d_{Rn}^i = d_{Rn}^i(R^i, p_n^i)$$

10 represents the rotational part of the shift and  $d_{Tn}^i$  represents the translational part. When

there is zero rotation,  $d_{Tn}^i = d_{Tn}^i$ . One can rewrite  $d_{Rn}^i = R^i * p_{Tn} - p_{Tn}$ , where

$p_{Tn} \equiv R^{-1} * p_n^i = p_n + d_{Tn}^i$ . We assume small translations and small residual rotations.

Then  $p_T \approx p_n + O(Z^{-1})$ ,

$$d_{Tn}^i \approx Z_n^{-1}(T_z^i p_n - [T^i]_2) + O(Z^{-2} \tau^2), \quad (2)$$

15  $d_{Rn}^i \approx \omega_x^i r^{(1)}(p_n) + \omega_y^i r^{(2)}(p_n) + \omega_z^i r^{(3)}(p_n) + O(\omega^2, \omega Z^{-1} \tau)$ , where  $Z^{-1}, \tau, \omega$  represent the

average sizes of the  $Z_i^{-1}$ , the translations and the residual rotations in radians. From "A

Linear Solution for Multiframe Structure from Motion", J. Oliensis, IUW 1225-1231,

1994 we get,  $\omega \approx Z^{-1} \tau$ .

Then using the brightness constancy equation (BCE),

$$20 \quad \Delta I_n^i + \nabla I_n \cdot d_n^i = 0, \quad (3)$$

which holds up to corrections of  $O(Z^{-2}\tau^2, \omega^2, \eta)$  where  $\eta$  gives the typical size of the noise in  $I_n^i$ . The brightness constancy equation and (2) imply that

$$-\Delta I_n^i \approx Z_n^{-1} (\nabla I_n \cdot \mathbf{p}_n T_z^i - \nabla I_n [T^i]_2) + \nabla I_n \cdot (\omega_x^i \mathbf{r}_n^{(1)} + \omega_y^i \mathbf{r}_n^{(2)} + \omega_z^i \mathbf{r}_n^{(3)}).$$

Then we define the three length- $N_p$  translational flow vectors as

$$\Phi_x \equiv -\{Z^{-1} I_x\}, \Phi_y \equiv -\{Z^{-1} I_y\}, \Phi_z \equiv -\{Z^{-1} (\nabla I_z \cdot \mathbf{p})\},$$

and also define the three length- $N_p$  rotational flow vectors as

$$\Psi_x \equiv \{\nabla I \cdot \mathbf{r}^{(1)}(\mathbf{p})\}, \Psi_y \equiv \{\nabla I \cdot \mathbf{r}^{(2)}(\mathbf{p})\}, \Psi_z \equiv \{\nabla I \cdot \mathbf{r}^{(3)}(\mathbf{p})\}.$$

Then let  $\Phi \equiv [\Phi_x \ \Phi_y \ \Phi_z]$  and  $\Psi \equiv [\Psi_x \ \Psi_y \ \Psi_z]$ . Then  $-\Delta \approx \bar{T}\Phi^T + W\Psi^T$ .

Then we define H as a  $(N_p - 3) \times N_p$  matrix that annihilates the three

- 10 vectors  $\Psi_x, \Psi_y, \Psi_z$  and satisfies  $HH^T = 1_{N_p-3}$ , where  $1_{N_p-3}$  is the identity matrix. One can then compute H, and products involving H, in  $O(N_p)$  using Householder matrices, which are described in "A Linear Solution for Multiframe Structure from Motion", J. Oliensis, IUW 1225-1231, 1994, and "A Multi-frame Structure from Motion Algorithm under Perspective Projection" J. Oliensis, IJCV 34:2/3, 163-192, 1999, and Workshop on
- 15 visual Scenes, 77-84, 1995. It then follows that

$$-\Delta H^T \approx \bar{T}\Phi^T H^T \quad (4)$$

up to  $O(Z^{-2}\tau^2, \omega^2, \omega Z^{-1}\tau, \eta)$ . In practice, we use equation (4) above left-multiplied by

$$C^{-1/2}, \text{ with } \Delta_{CH} \equiv C^{-1/2} \Delta H^T. \text{ Multiplying by } C^{-1/2} \text{ reduces the bias due to singling out}$$

the reference image for special treatment, a process described in "A Multi-frame

- 20 Structure from Motion Algorithm under Perspective Projection" which is referenced above.

Equation (4) relates the data-matrix, on the left, to the translations and structure, on the right. Multiplying by H has eliminated the rotational effects up to second order. These second order corrections include corrections of  $O(\omega\eta)$ , caused by errors in the measured  $\nabla I$  we use to define H. For small translations,  $O(Z^{-1}\tau) \sim O(\omega)$  as described in

5 "Rigorous Bounds for Two-Frame Structure from Motion," J. Oliensis IUW, 1225-1231, 1994 so all the corrections in equation (4) have similar

sizes:  $O(Z^{-2}\tau^2) \sim O(\omega Z^{-1}\tau) \sim O(\omega^2)$ . Therefore, multiplying by H was crucial to reduce the rotational corrections to the same order as the translational corrections.

## 10 Linear-Motion Algorithm

The basic algorithm of the present invention for cases of linear camera motion is more particularly described as follows.

0. Recover rotations and warp all images  $I^1 \dots I^{N_1-1}$  toward the reference image  $I^0$ , while neglecting the translations. Let the image displacements  $d_n^i$  now
- 15 refer to the unrotated images.
1. Compute H and  $\Delta_{CH}$ . Using the singular value decomposition, compute the best rank-1 factorization of  $-\Delta_{CH} \approx M^{(1)}S^{(1)T}$ , where  $M^{(1)}, S^{(1)}$  are vectors. If the leading singular value of  $-\Delta_{CH}$  is much larger than the rest, this confirms that the motion is approximately linear and that the signal dominates the noise so that
- 20 the algorithm has enough information to proceed.  $C^{1/2}M^{(1)}$  gives the translation magnitudes up to an overall multiplicative scale.



2. Divide the image into small smoothing windows and take  $Z_n^{-1}$  as constant within each window. List the pixels so that those in the  $k$ -th smoothing window have sequential indices  $\eta_k, (\eta_k + 1), \dots, (\eta_{k+1} - 1)$ . Then compute a  $N_p \times N_p$  projection matrix  $P_\Omega$  which is block diagonal with zero entries between different smoothing windows, and which annihilates the vectors  $\{\nabla I \cdot \mathbf{p}\}, \{I_x\}$ , and  $\{I_y\}$ .

Then solve the overconstrained system of equations

$$P_\Omega (H^T S^{(1)} - \Psi w) = 0 \quad (5)$$

for the 3-vector  $w$ .

To complete the method of the Linear-Motion Algorithm, compute a  $N_p \times N_p$

projection matrix  $P_T$ , which is block diagonal with zero entries between different smoothing windows and annihilates  $(H^T S^{(1)}) - \Psi w$  where  $w$  is the vector recovered previously. Then solve for the direction of translation  $\hat{T}$  via

$$P_{\hat{T}} (-\hat{T}_x \{I_x\} - \hat{T}_y \{I_y\} + \hat{T}_z \{p \cdot \nabla I\}) = 0 \quad (6)$$

Finally, recover  $Z_n$  via

$$(H^T S^{(1)})_n - [\Psi w]_n = Z_n^{-1} (\hat{T}_z p_n - [\hat{T}]_2) \cdot \nabla I_n \quad (7)$$

### Linear-Motion Algorithm Analysis

Step 2.

From (4),  $S^{(1)} \sim H \{Z^{-1} (\hat{T}_z p - [\hat{T}]_2) \cdot \nabla I\}$ . Then since  $H^T H$  is a projection matrix annihilating the  $\Psi_x, \Psi_y, \Psi_z$  it follows that

$$H^T S^{(1)} \sim \{Z^{-1}(\hat{T}_z p - [\hat{T}_2]) \cdot \nabla I\} + \Psi w \quad (8)$$

for some  $w$ . Since the matrix  $P_\Omega$  annihilates the first term on the right hand side of (8),

we get (5).  $P_\Omega$  and its products can be computed in  $O(N_p)$ . Solving (5) for  $w$  neglects

the constraints that  $\hat{T}$  is the same in each smoothing window, a total of  $2(N_w - 1)$

5 constraints, where  $N_w$  is the number of windows. Then applying  $P_{\hat{T}}$  to (8) gives (6).

Because we omitted  $2(N_w - 1)$  constraints, Step 2 gives a suboptimal estimate of

$\hat{T}$  and  $Z_n^{-1}$ . As before, one can base Step 2 on a multi-frame reestimate of  $I^0$  and as

before the caveat that if the original noise in  $I^0$  is less than the recomputed  $I^0$ , one should use  $I^0$  directly.

10 The linear-motion algorithm extends to deal with a camera translating on a plane or in all 3D directions. The number  $N_L$  of large singular values  $\Delta_{CH}$  determines the dimensionality of the motion, e.g., planar motion corresponds to  $N_L = 2$ . For each large singular value, the corresponding singular vector gives rise to an equation similar to (8), which can be solved as before for  $\hat{T}$ , where each singular vector yields a different  $\hat{T}$ .

15 One recovers the  $Z_n^{-1}$  from  $N_L$  equations of the form of (7).

### **Implementation**

It will be apparent to those skilled in the art that the methods of the present invention disclosed herein may be embodied and performed completely by software  
20 contained in an appropriate storage medium for controlling a computer.

Referring to Fig. 1, which illustrates in block-diagram form a computer hardware system incorporating the invention. As indicated therein, the system includes a video

source 101 , whose output is digitized into a pixel map by a digitizer 102. The digitized video frames are then sent in electronic form via a system bus 103 to a storage device 104 for access by the main system memory during usage. During usage the operation of the system is controlled by a central-processing unit, (CPU) 105 which controls the access to the digitized pixel map and the invention. The computer hardware system will include those standard components well-known to those skilled in the art for accessing and displaying data and graphics, such as a monitor, 106 and graphics board 107.

The user interacts with the system by way of a keyboard 108 and or a mouse 109 or other position-sensing device such as a track ball, which can be used to select items on the screen or direct functions of the system.

The execution of the key tasks associated with the present invention is directed by instructions stored in the main memory of the system, which is controlled by the CPU. The CPU can access the main memory and perform the steps necessary to carry out the method of the present invention in accordance with instructions stored that govern CPU operation. Specifically, the CPU, in accordance with the input of a user will access the stored digitized video and in accordance with the instructions embodied in the present invention will analyze the selected video images in order to extract the 3D structure information from the associated digitized pixel maps.

Referring now to Fig. 2 the method of the present invention will be described in relation to the block diagram. A first image in a sequence is taken by a camera at a reference perspective and one or more successive images are taken by moving the camera along a substantially linear plane to one or more successive different perspectives in step 201. The images are then digitized 202 for analysis of the 3D image content, i.e. image

intensities. From the digitized 3D image content, determining image data shifts for each successive image 203 with respect to the reference image; the shifts being derived from the camera translation and/or rotation from the reference perspective to the successive different perspectives.

5        Then incorporating the image data shifts for each image, constructing a shift data matrix 204. The shift data matrix is then used to calculate two rank-1 factorizations from the shift data matrix using SVD, one rank-1 factorization being a vector corresponding the 3D structure and the other rank-1 factorization being a vector corresponding the camera motion 205. The successive images are divided into smoothing windows 206 and  
10    the camera motion is recovered from the factorization vectors between the smoothing windows by solving a linear equation 207. Finally, the 3D structure is recovered by solving a linear equation using the recovered camera motion 208.

While there has been shown and described what is considered to be preferred  
embodiments of the invention, it will, of course, be understood that various modifications  
15    and changes in the form or detail could readily be made without departing from the spirit of the invention. It is therefore intended that the invention be not limited to the exact forms described and illustrated, but should be constructed to cover all modifications that may fall within the scope of the appended claims.

**WHAT IS CLAIMED IS:**

1. A method for recovering 3D scene structure and camera motion from image data obtained from a multi-image sequence, wherein a reference image of the sequence is  
5 taken by a camera at a reference perspective and one or more successive images of the sequence are taken at one or more successive different perspectives by translating and/or rotating the camera, the method comprising the steps of:

(a) determining image data shifts for each successive image with respect to the reference image; the shifts being derived from the camera translation and/or rotation from  
10 the reference perspective to the successive different perspectives;

(b) constructing a shift data matrix that incorporates the image data shifts for each image;

(c) calculating a rank-1 factorization from the shift data matrix using SVD, with one of the rank-1 factors being a vector corresponding to the 3D structure and the other  
15 rank-1 factor being a vector corresponding to the size of the camera motions;

(d) dividing the successive images into smoothing windows;

(e) recovering the direction of camera motion from the first vector corresponding to the 3D structure by solving a linear equation; and

(f) recovering the 3D structure by solving a linear equation using the recovered  
20 camera motion.

2. The method of claim 1, wherein, step (e) includes:

computing a first projection matrix;

recovering camera rotation vectors from the shift data matrix, and the first projection matrix;

computing a second projection matrix; and

recovering the direction of camera translation using the shift data matrix, the reference image, the second projection matrix and the recovered camera rotation vectors.

3. The method of claim 2, wherein step (f) includes recovering the 3D structure from the shift data matrix, the reference image, the recovered camera rotation vectors and the recovered direction of translation vectors.

4. The method of claim 1, further including preliminary steps of :  
recovering the rotations of the camera between each successive image; and  
warping all images in the sequence toward the reference image, while neglecting the translations.

5. The method of claim 1, wherein step (b) comprises:  
computing H and  $\Delta_{CH}$ , where H is a  $(N_p - 3) \times N_p$  matrix defined so that  $HH^T$  is the identity matrix and H annihilates the three vectors  $\Psi_x, \Psi_y, \Psi_z$  where the three vectors are computed from the reference image as

$$\Psi_x \equiv \{\nabla I \cdot \mathbf{r}^{(1)}(\mathbf{p})\}, \Psi_y \equiv \{\nabla I \cdot \mathbf{r}^{(2)}(\mathbf{p})\}, \Psi_z \equiv \{\nabla I \cdot \mathbf{r}^{(3)}(\mathbf{p})\} \text{ where}$$

$r^{(1)}(x, y), r^{(2)}(x, y), r^{(3)}(x, y)$  are defined by  $[r^{(1)}, r^{(2)}, r^{(3)}] \equiv \left[ \begin{pmatrix} -xy \\ -(1+y^2) \end{pmatrix}, \begin{pmatrix} 1+x^2 \\ xy \end{pmatrix}, \begin{pmatrix} -y \\ x \end{pmatrix} \right]$

and  $\Delta$  is a shift data matrix, that gives the difference in intensities between each

successive image and the reference image and is a  $(N_I - 1) \times N_p$  matrix with entries  $\Delta I_n^i$

, where  $\Delta I_n^i$  is the change in (smoothed) intensity with respect to the reference image,

5 and with no smoothing  $\Delta I_n^i = I_n^i - I_n^0$ , where  $N_I$  is the number of images,  $N_p$  is the

number of pixels, and where  $I^i$  denotes the  $i$ -th image, with  $i=0, 1, \dots, N_I - 1$ , and where

$I_n^i = I^i(p_n)$  denotes the image intensity at the  $n$ -th pixel position in  $I^i$ , where  $I^0$  is the

reference image, where  $x$  and  $y$  are the image coordinates of the pixel position and

$p = (x, y)$  and where  $\Delta_{CH} \equiv C^{-1/2} \Delta H^T$  where  $C$  is a constant matrix with and where the

10 notation  $\{V\}$  used to denote a vector with elements given by the  $V^a$ .

6. The method of claim 1, wherein step (c) comprises:

computing a rank-1 factorization of  $-\Delta_{CH} \approx M^{(1)} S^{(1)T}$  where  $M^{(1)}, S^{(1)T}$  are

vectors corresponding to the motion and structure respectively.

15

7. The method of claim 1, wherein step (c) comprises:

computing a rank-3 factorization of  $-\Delta_{CH} \approx \sum_{a=1}^3 M^{(a)} S^{(a)T}$  where  $M^{(a)}, S^{(a)T}$  are

vectors corresponding to the motion and structure respectively;

setting  $Z_n^{-1}$  as constant within each window, where  $Z$  is the depth from the

20 camera to a 3D scene along the cameras optical axis;

listing each of the pixels so that those in the  $k$ -th smoothing window have sequential indices  $n_k, (n_k+1), \dots (n_{k+1} - 1)$ ;

computing a first projection matrix by computing a  $N_p \times N_p$  projection matrix  $P_\Omega$  which is block diagonal with zero entries between different smoothing windows, and

5 which annihilates the vectors  $\{\nabla I \cdot p\}, \{I_x\}$  and  $\{I_y\}$  where  $\{\nabla I\}$  is a vector containing the gradient of the intensity at each pixel, and  $I_x$  and  $I_y$  are the gradients of the intensity in the reference image in the  $x$  and  $y$  directions;

recovering the three camera rotation vectors includes solving the following equations

10 
$$P_\Omega (H^T S^{(a)} - \Psi w^{(a)}) = 0 \text{ for the 3-vector } w^{(a)};$$

computing a second projection matrix includes computing a  $N_p \times N_p$  projection matrix  $P_T^{(a)}$ , which is block diagonal with zero entries between different smoothing windows and annihilates  $(H^T S^{(a)}) - \Psi w^{(a)}$  where  $w^{(a)}$  is the vector recovered previously;

15 recovering the directions of camera translations by solving for the directions of translation  $\hat{T}^{(a)}$  via

$$P_{\hat{T}}^{(a)} (-\hat{T}_x^{(a)} \{I_x\} - \hat{T}_y^{(a)} \{I_y\} + \hat{T}_z^{(a)} \{p \cdot \nabla I\}) = 0 \text{ and};$$

recovering  $Z_n$  via

$$(H^T S^{(a)})_n - [\Psi w^{(a)}]_n = \tau^{(a)} (Z_n^{-1} (\hat{T}_z^{(a)} p_n - [\hat{T}^{(a)}]_2) \cdot \nabla I_n) \text{ where } [\hat{T}^{(a)}]_2 \text{ represents } \hat{T}_x^{(a)} \hat{T}_y^{(a)}$$

20 the  $x$  and  $y$  component of the translation direction, and where  $\tau^{(a)}$  are constants.

8. The method of claim 1, wherein step (d) comprises:



listing each of the pixels so that those in the  $k$ -th smoothing window have sequential indices  $n_k, (n_k+1), \dots (n_{k+1}-1)$ .

9. The method of claim 2, wherein the step of computing a first projection matrix includes computing a  $N_p \times N_p$  projection matrix  $P_\Omega$  which is block diagonal with zero entries between different smoothing windows, and which annihilates the vectors  $\{\nabla I \cdot p\}, \{I_x\}$  and  $\{I_y\}$  where  $\{\nabla I\}$  is a vector containing the gradient of the intensity at each pixel, and  $I_x$  and  $I_y$  are the gradients of the intensity in the reference image in the  $x$  and  $y$  directions.

15  $P_{\Omega}(H^T S^{(1)} - \Psi_w) = 0$  for the 3-vector  $w$ .

20

$$\mathbf{P}_{\hat{T}}\left(-\hat{T}_x\{I_x\}-\hat{T}_y\{I_y\}+\hat{T}_z\{\mathbf{p}\cdot\nabla I\}\right)=0.$$

$(H^T S^{(1)})_n - [\Psi w]_n = Z_n^{-1} (\hat{T}_z \mathbf{p}_n - [\hat{\mathbf{T}}]_2) \cdot \nabla I_n$  where  $[\hat{\mathbf{T}}]_2$  represents  $\hat{\mathbf{T}}_x \hat{\mathbf{T}}_y$  the x and y component of the translation direction.

**A METHOD FOR RECOVERING 3D SCENE STRUCTURE AND CAMERA  
MOTION DIRECTLY FROM IMAGE INTENSITIES**

5

**ABSTRACT OF THE INVENTION**

The present invention is directed to a method for recovering 3D scene  
10 structure and camera motion from image data obtained from a multi-image sequence,  
wherein a reference image of the sequence is taken by a camera at a reference perspective  
and one or more successive images of the sequence are taken at one or more successive  
different perspectives by translating and/or rotating the camera, the method comprising  
The steps of determining image data shifts for each successive image with respect to the  
15 reference image; the shifts being derived from the camera translation and/or rotation from  
the reference perspective to the successive different perspectives;  
constructing a shift data matrix that incorporates the image data shifts for each image;  
calculating a rank-1 factorizations from the shift data matrix using SVD, with one of the  
rank-1 factors being a vector corresponding to the 3D structure and the other rank-1  
20 factor being a vector corresponding to the size of the camera motions; dividing the  
successive images into smoothing windows; recovering the direction of camera motion  
from the first vector corresponding to the 3D structure by solving a linear equation; and  
recovering the 3D structure by solving a linear equation using the recovered camera  
motion. In accordance with the present invention, the method includes computing a first  
25 projection matrix; recovering camera rotation vectors from the shift data matrix, and the  
first projection matrix; computing a second projection matrix; and recovering the  
direction of camera translation using the shift data matrix, the reference image, the

second projection matrix and the recovered camera rotation vectors. In addition the method includes recovering the 3D structure from the shift data matrix, the reference image, the recovered camera rotation vectors and the recovered direction of translation vectors.

5

**REPORT OF THE**

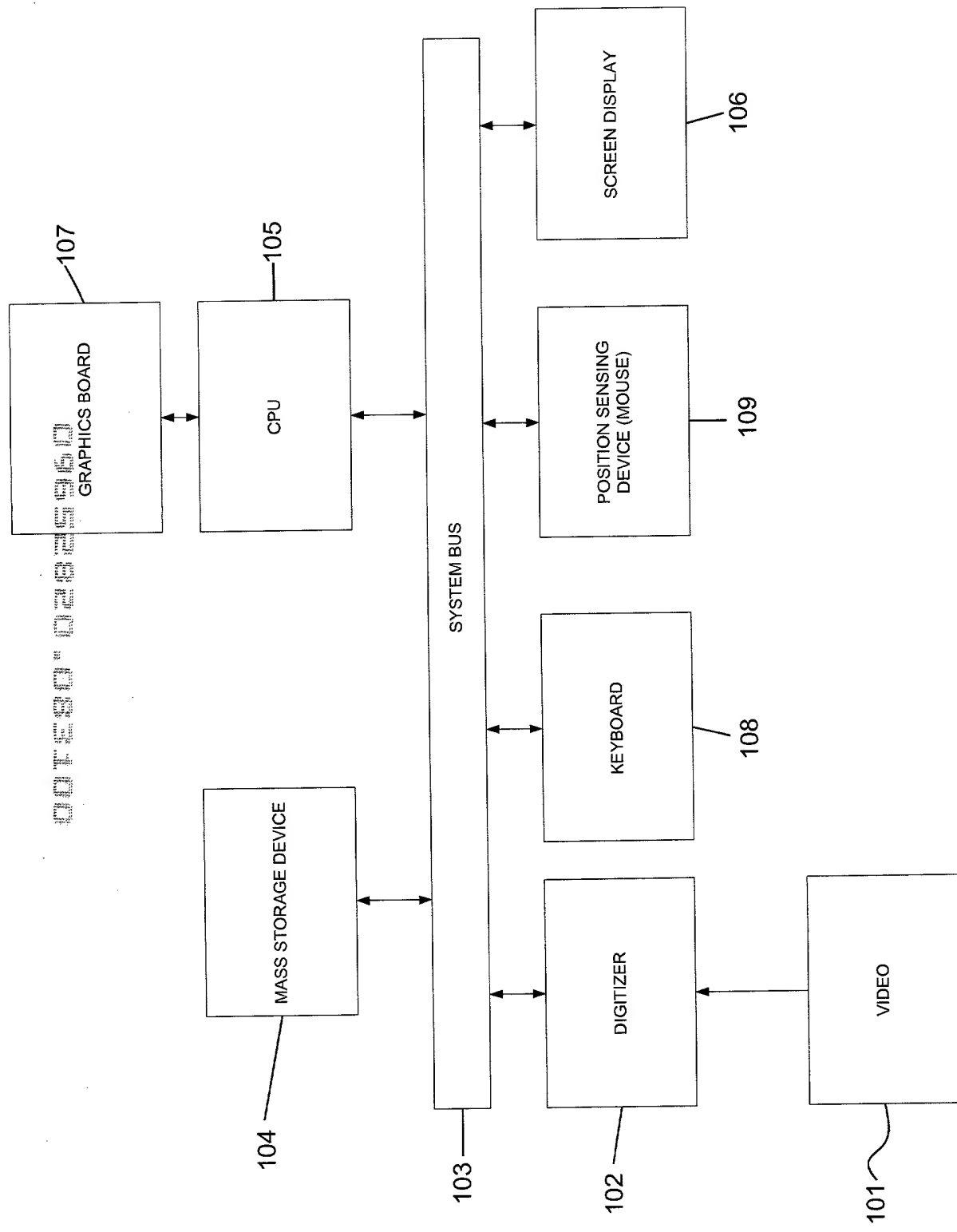


Fig. 1

00000000-00000000

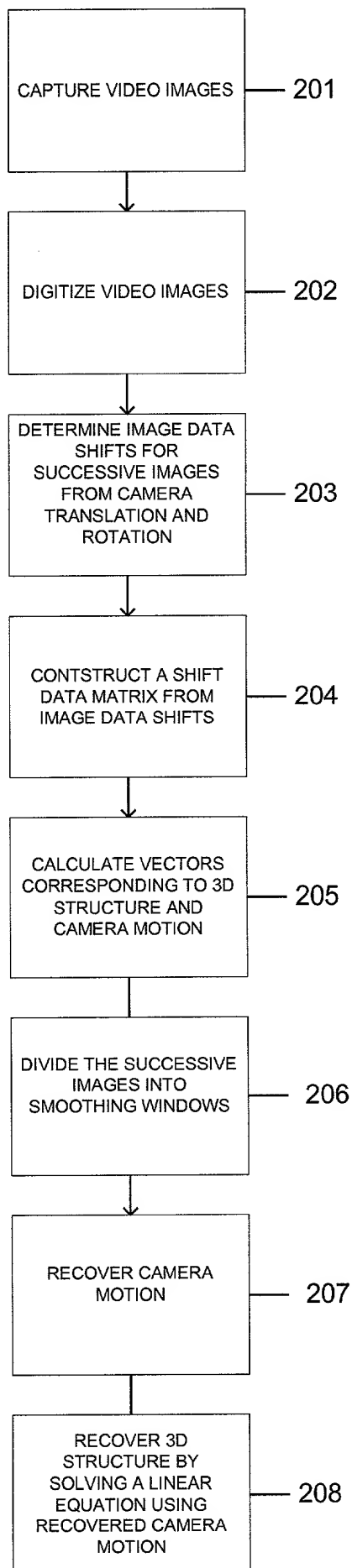


Fig. 2

Docket No.

13725 (NECI-1086)

## Declaration and Power of Attorney For Patent Application

### English Language Declaration

As a below named inventor, I hereby declare that:

My residence, post office address and citizenship are as stated below next to my name.

I believe I am the original, first and sole inventor (if only one name is listed below) or an original, first and joint inventor (if plural names are listed below) of the subject matter which is claimed and for which a patent is sought on the invention entitled

**A METHOD FOR RECOVERING 3D SCENE STRUCTURE AND CAMERA MOTION DIRECTLY FROM IMAGE INTENSITIES**

the specification of which

(check one)

☒ is attached hereto.

☐ was filed on \_\_\_\_\_ as United States Application No. or PCT International Application Number \_\_\_\_\_ and was amended on \_\_\_\_\_ (if applicable)

I hereby state that I have reviewed and understand the contents of the above identified specification, including the claims, as amended by any amendment referred to above.

I acknowledge the duty to disclose to the United States Patent and Trademark Office all information known to me to be material to patentability as defined in Title 37, Code of Federal Regulations, Section 1.56.

I hereby claim foreign priority benefits under Title 35, United States Code, Section 119(a)-(d) or Section 365(b) of any foreign application(s) for patent or inventor's certificate, or Section 365(a) of any PCT International application which designated at least one country other than the United States, listed below and have also identified below, by checking the box, any foreign application for patent or inventor's certificate or PCT International application having a filing date before that of the application on which priority is claimed.

Prior Foreign Application(s)

Priority Not Claimed

(Number)

(Country)

(Day/Month/Year Filed)

☐

(Number)

(Country)

(Day/Month/Year Filed)

☐

(Number)

(Country)

(Day/Month/Year Filed)

☐

I hereby claim the benefit under 35 U.S.C. Section 119(e) of any United States provisional

_____ (Application Serial No.)	_____ (Filing Date)
-----------------------------------	------------------------

_____ (Application Serial No.)	_____ (Filing Date)
-----------------------------------	------------------------

_____ (Application Serial No.)	_____ (Filing Date)
-----------------------------------	------------------------

I hereby claim the benefit under 35 U. S. C. Section 120 of any United States application(s), or Section 365(c) of any PCT International application designating the United States, listed below and, insofar as the subject matter of each of the claims of this application is not disclosed in the prior United States or PCT International application in the manner provided by the first paragraph of 35 U.S.C. Section 112, I acknowledge the duty to disclose to the United States Patent and Trademark Office all information known to me to be material to patentability as defined in Title 37, C. F. R., Section 1.56 which became available between the filing date of the prior application and the national or PCT International filing date of this application:

_____ (Application Serial No.)	_____ (Filing Date)	_____ (Status) (patented, pending, abandoned)
-----------------------------------	------------------------	---

_____ (Application Serial No.)	_____ (Filing Date)	_____ (Status) (patented, pending, abandoned)
-----------------------------------	------------------------	---

_____ (Application Serial No.)	_____ (Filing Date)	_____ (Status) (patented, pending, abandoned)
-----------------------------------	------------------------	---

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.



POWER OF ATTORNEY: As a named inventor, I hereby appoint the following attorney(s) and/or agent(s) to prosecute this application and transact all business in the Patent and Trademark Office connected therewith. (list name and registration number)

Stephen D. Murphy; Reg. No. 22,002  
William C. Roch; Reg. No. 24,972  
Frank S. DiGiglio; Reg. No. 31,346

Paul J. Esatto, Jr.; Reg. No. 30,749  
Mark J. Cohen; Reg. No. 32,211  
Philip J. Felg; Reg. No. 27,328  
Andrew G. Izzwan; Reg. No. 40,028

Send Correspondence to: Paul J. Esatto, Jr., Esq.  
Scully, Scott, Murphy & Plesser  
400 Garden City Plaza  
Garden City, New York 11530

Direct Telephone Calls to: (name and telephone number)  
Paul J. Esatto, Jr. Esq. at (516) 742-4343

Full name of sole or first inventor	
John Ollensis	
Sole or first inventor's signature	Date
John Ollensis	Aug 31 2000
Residence	
246 W. 71st Street, New York, New York 10023	
Citizenship	
USA	
Post Office Address	
same as above	

Full name of second inventor, if any	
Second inventor's signature	Date
Residence	
Citizenship	
Post Office Address	